WHAT IS CLAIMED IS:

1. A method for selecting from a plurality of different siRNAs one or more siRNAs for silencing a target gene in an organism, each of said plurality of different siRNAs targeting a different target sequence in a transcript of said target gene, said method comprising

5   (a) ranking said plurality of different siRNAs according to positional base compositions of a corresponding targeted sequence motifs in said transcript, wherein each said targeted sequence motif comprises at least a portion of the target sequence of the corresponding siRNA and/or a second sequence in a sequence region flanking said target sequence; and

10   (b) selecting one or more siRNAs from said ranked siRNAs.

2. The method of claim 1, wherein each said sequence motif comprises said target sequence of said targeting siRNA.

3. The method of claim 2, wherein said ranking step is carried out by (a1) determining a score for each said different siRNA, wherein said score is calculated using a position-
15   specific score matrix; and (a2) ranking said plurality of different siRNAs according to said score.

4. The method of claim 3, wherein each said sequence motif is a nucleotide sequence of L nucleotides, L being an integer, and wherein said position-specific score matrix is $\{\log(e_{ij}/p_{ij})\}$, where e_{ij} is the weight of nucleotide i at position j, p_{ij} is the weight of nucleotide
20   i at position j in a random sequence, and $i = G, C, A, U(T)$, $j = 1, ..., L$.

5. The method of claim 3, wherein each said sequence motif is a nucleotide sequence of L nucleotides, L being an integer, and wherein said position-specific score matrix is $\{\log(e_{ij}/p_{ij})\}$, where e_{ij} is the weight of nucleotide i at position j, p_{ij} is the weight of nucleotide i at position j in a random sequence, and $i = G$ or $C, A, U(T)$, $j = 1, ..., L$.

25   6. The method of claim 5, wherein said score for each said siRNA is calculated according to equation

$$Score = \sum_{i=1}^{L} \ln(e_i / p_i)$$

wherein said e_t and p_t are respectively weights of the nucleotide at position t in said sequence motif as determined according to said position-specific score matrix and in a random sequence.

7. The method of claim 6, wherein each said sequence motif comprises said target sequence of said targeting siRNA and at least one flanking sequence.

8. The method of claim 7, wherein each said sequence motif comprises said target sequence of said targeting siRNA and a 5' flanking sequence and a 3' flanking sequence.

9. The method of claim 8, wherein said 5' flanking sequence and said 3' flanking sequence are each a sequence of D nucleotides, D being an integer.

10. The method of claim 9, wherein each said target sequence is a sequence of 19 nucleotides, and each said 5' flanking sequence and 3' flanking sequence are a sequence of 10 nucleotides.

11. The method of claim 8, wherein each said target sequence is a sequence of 19 nucleotides, and each said 5' flanking sequence and 3' flanking sequence are a sequence of 50 nucleotides.

12. The method of claim 10, wherein said one or more siRNAs consist of at least 3 siRNAs.

13. The method of claim 12, further comprising a step of de-overlapping, said step of de-overlapping comprising selecting a plurality of siRNAs among said at least 3 siRNAs such that siRNAs in said plurality are sufficiently different in a sequence diversity measure.

14. The method of claim 13, wherein said diversity measure is a quantifiable measure, and said selecting in said de-overlapping step comprises selecting siRNAs having a difference in said sequence diversity measure between different selected siRNAs above a given threshold.

15. The method of claim 14, wherein said sequence diversity measure is the overall GC content of said siRNAs.

16. The method of claim 15, wherein said given threshold is 5%.

17. The method of claim 14, wherein said sequence diversity measure is the distance between siRNAs along the length of the transcript sequence.

18. The method of claim 17, wherein said threshold is 100 nucleotides.

19. The method of claim 14, wherein said sequence diversity measure is the identity of the leading dimer of said siRNAs, wherein each of the 16 possible leading dimers is assigned a score of 1-16, respectively.

20. The method of claim 19, wherein said threshold is 0.5.

21. The method of claim 1, further comprising a step of selecting one or more siRNAs based on silencing specificity, said step of selecting based on silencing specificity comprising, (i) for each of said plurality of siRNAs, predicting off-target genes of said siRNA from among a plurality of genes, wherein said off-target genes are genes other than said target gene and are directly silenced by said siRNA; (ii) ranking said plurality of siRNAs according to the number of off-target genes; and (iii) selecting one or more siRNAs for which said number of off-target genes is below a given threshold.

22. The method of claim 21, wherein said predicting comprises (i1) evaluating sequence of each of said plurality of genes based on a predetermined siRNA sequence match pattern; and (i2) predicting said gene as an off-target gene if said gene comprise a sequence that matches said siRNA based on said sequence match pattern.

23. The method of claim 22, wherein said step of evaluating comprises identifying an alignment of said siRNA to a sequence in a gene by a low stringency FastA alignment.

24. The method of claim 23, wherein each said siRNA has L nucleotides in its duplex region, and wherein said match pattern is represented by a position match position-specific score matrix (pmPSSM), said position match position-specific score matrix consisting of weights of different positions in an siRNA to match transcript sequence positions in an off-target transcript $\{P_j\}$, where $j = 1, ..., L$, P_j is the weight of a match at position j.

25. The method of claim 24, wherein said step (i1) comprises calculating a position match score pmScore according to equation

$$pmScore = \sum_{i=1}^{L} \ln(E_i / 0.25)$$

where $E_i = P_i$ if position i is a match and $E_i = (1-P_i)/3$ if position i is a mismatch; and said step (i2) comprises predicting said gene as an off-target gene if said position match score is greater than a given threshold.

26. The method of claim 25, wherein L is 19, and wherein said pmPSSM is given by Table I.

27. The method of claim 26, wherein said plurality of genes comprises all known unique genes of said organism other than said target gene.

28. The method of claim 10, wherein said position-specific score matrix (PSSM) is obtained by a method comprising

(aa) identifying a plurality of N siRNAs consisting of siRNAs having 19-nucleotide duplex region and having a silencing efficacy above a chosen threshold;

(bb) identifying for each said siRNA a functional sequence motif, said functional sequence motif comprising a 19-nucleotide target sequence of said siRNA and a 10-nucleotide 5' flanking sequence and a 10-nucleotide 3' flanking sequence;

(cc) calculating a frequency matrix $\{f_{ij}\}$, where $i = G, C, A, U(T); j = 1, 2, ..., L$, and where f_{ij} is the frequency of the ith nucleotide at the jth position, based on said siRNAs functional sequence motifs according to equation

$$f_{ij} = \sum_{k=1}^{N} \delta_{ik}(j),$$

where $\delta_{ik}(j) = \begin{cases} 1, \text{if } k = i \\ 0, \text{if } k \neq i \end{cases}$, and

(dd) determining said PSSM by calculating e_{ij} according to equation

$$e_{ij} = \frac{f_{ij}}{N}.$$

29. The method of claim 28, wherein said plurality of N siRNAs target a plurality of different genes having different transcript abundances in a cell.

30. The method of claim 29, wherein said step (b) is carried out by selecting one or more siRNAs having the highest scores.

31. The method of claim 29, wherein said step (b) is carried out by selecting one or more siRNAs having a score closest to a predetermined value, wherein said predetermined value is the score value corresponding to the maximum median silencing efficacy of a plurality of siRNA sequence motifs.

32. The method of claim 31, wherein said plurality of siRNA sequence motifs are sequence motifs in transcript having abundance level of less than about 3-5 copies per cell.

33. The method of claim 29, wherein said step (b) is carried out by selecting one or more siRNAs having a score within a predetermined range, wherein said predetermined range is a score range corresponding to a plurality of siRNAs sequence motifs having a given level of silencing efficacy.

34. The method of claim 33, wherein said silencing efficacy is above 50%, 75%, or 90% at an siRNA dose of about 100nM.

35. The method of claim 34, wherein said plurality of siRNA sequence motifs are sequence motifs in transcript having abundance level of less than about 3-5 copies per cell.

36. The method of any one of claims 28-35, wherein said plurality of N siRNAs comprises at least 10, 50, 100, 200, or 500 different siRNAs.

37. The method of any one of claims 5-11, wherein said position-specific score matrix (PSSM) is obtained by a method comprising

(aa) initializing said PSSM with random weights;

(bb) selecting randomly a weight w_{ij} obtained in (aa);

(cc) changing the value of said selected weight to generate a test psPSSM comprising said selected weight having said changed value;

(dd) calculating a score for each of a plurality of siRNAs functional sequence motifs using said test PSSM according to equation

$$Score = \sum_{k=1}^{L} \ln(w_k / p_k)$$

wherein said w_k and p_k are respectively weights of a nucleotide at position k in said functional sequence motif and in a random sequence;

(ee) calculating correlation of said score and a metric of a characteristic of an siRNA among said plurality of siRNAs functional sequence motifs;

(ff) repeating steps (cc)-(ee) for a plurality of different values of said selected weight in a given range and retain the value that corresponds to the best correlation for said selected weight; and

(gg) repeating steps (bb)-(ff) for a chosen number of times; thereby determining said PSSM.

38. The method of claim 37, further comprising selecting said plurality of siRNA functional sequence motifs by a method comprising

(i) identifying a plurality of siRNAs consisting of siRNAs having different values in said metric;

(ii) identifying a plurality of siRNA functional sequence motifs each corresponding to an siRNA in said plurality of siRNAs.

39. The method of claim 38, wherein said characteristic is silencing efficacy.

40. The method of claim 39, wherein said plurality of N siRNAs target a plurality of different genes having different transcript abundances in a cell.

41. The method of claim 40, wherein said step (b) is carried out by selecting one or more siRNAs having the highest scores.

42. The method of claim 40, wherein said step (b) is carried out by selecting one or more siRNAs having a score closest to a predetermined value, wherein said predetermined

value is the score value corresponding to the maximum median silencing efficacy of a plurality of siRNA sequence motifs.

43. The method of claim 42, wherein said plurality of siRNA sequence motifs are sequence motifs in transcript having abundance level of less than about 3-5 copies per cell.

5   44. The method of claim 40, wherein said step (bb) is carried out by selecting one or more siRNAs having a score within a predetermined range, wherein said predetermined range is a score range corresponding to a plurality of siRNAs sequence motifs having a given level of silencing efficacy.

45. The method of claim 44, wherein said silencing efficacy is above 50%, 75%, or 10   90% at an siRNA dose of about 100nM.

46. The method of claim 45, wherein said plurality of siRNA sequence motifs are sequence motifs in transcript having abundance level of less than about 3-5 copies per cell.

47. The method of any one of claims 39-46, wherein said plurality of N siRNAs comprises at least 10, 50, 100, 200, or 500 different siRNAs.

15   48. The method of claim 37, wherein said position-specific score matrix (PSSM) comprises w_k, $k = 1, ..., L$, w_k being a difference in probability of finding nucleotide G or C at sequence position k between a first type of siRNA and a second type of siRNA, and wherein said score for each said strand is calculated according to equation

$$Score = \sum_{k=1}^{L} w_k .$$

20   49. The method of claim 48, wherein said first type of siRNA consists of one or more siRNAs having silencing efficacy no less than a first threshold and said second type of siRNA consists of one or more siRNAs having silencing efficacy less than a second threshold.

50. The method of claim 49, wherein said difference in probability is described by a 25   sum of Gaussian curves, each of said Gaussian curves representing the difference in probability of finding a G or C at a different sequence position .

51. The method of claim 50, wherein said first and second threshold are both 75% at an siRNA dose of 100nM.

52. A method for selecting from a plurality of different siRNAs one or more siRNAs for silencing a target gene in an organism, each of said plurality of different siRNAs targeting a different target sequence in a transcript of said target gene, said method comprising

(a) ranking said plurality of different siRNAs according to positional base composition of reverse complement sequences of sense strands of said siRNAs; and

(b) selecting one or more siRNAs from said ranked siRNAs.

53. The method of claim 52, wherein said ranking step is carried out by (a1) determining a score for each said different siRNA, wherein said score is calculated using a position-specific score matrix; and (a2) ranking said plurality of different siRNAs according to said score.

54. The method of claim 53, wherein said siRNA has a nucleotide sequence of L nucleotides in its duplex region, L being an integer, wherein said position-specific score matrix comprises w_k, $k = 1, ..., L$, w_k being a difference in probability of finding nucleotide G or C at sequence position k between reverse complement of sense strand of a first type of siRNA and reverse complement of sense strand of a second type of siRNA, and wherein said score for each said reverse complement is calculated according to equation

$$Score = \sum_{k=1}^{L} w_k \; .$$

55. The method of claim 54, wherein said first type of siRNA consists of one or more siRNAs having silencing efficacy no less than a first threshold and said second type of siRNA consists of one or more siRNAs having silencing efficacy less than a second threshold.

56. The method of claim 55, wherein said difference in probability is described by a sum of Gaussian curves, each of said Gaussian curves representing the difference in probability of finding a G or C at a different sequence position .

57. The method of claim 56, wherein said first and second threshold are both 75% at an siRNA dose of 100nM.

58. A method for selecting from a plurality of different siRNAs one or more siRNAs for silencing a target gene in an organism, each of said plurality of different siRNAs targeting a different target sequence in a transcript of said target gene, said method comprising, (i) for each of said plurality of different siRNAs, predicting off-target genes of said siRNA from among a plurality of genes, wherein said off-target genes are genes other than said target gene and are directly silenced by said siRNA; (ii) ranking said plurality of different siRNAs according to the number of off-target genes; and (iii) selecting one or more siRNAs for which said number of off-target genes is below a given threshold.

59. The method of claim 58, wherein said predicting comprises (i1) evaluating sequence of each of said plurality of genes based on a predetermined siRNA sequence match pattern; and (i2) predicting said gene as an off-target gene if said gene comprise a sequence that matches said siRNA based on said sequence match pattern.

60. The method of claim 59, wherein each said siRNA has L nucleotides in its duplex region, and wherein said sequence match pattern is represented by a position match position-specific score matrix (pmPSSM), said position match position-specific score matrix consisting of weights of different positions in an siRNA to match transcript sequence positions in an off-target transcript $\{P_j\}$, where $j = 1, ..., L$, P_j is the weight of a match at position j.

61. The method of claim 60, wherein said step (i1) comprises calculating a position match score pmScore according to equation

$$pmScore = \sum_{i=1}^{L} \ln(E_i / 0.25)$$

where $E_i = P_i$ if position i is a match and $E_i = (1-P_i)/3$ if position i is a mismatch; and said step (i2) comprises predicting said gene as an off-target gene if said position match score is greater than a given threshold.

62. The method of claim 61, wherein L is 19, and wherein said pmPSSM is given by Table I.

63. The method of claim 62, wherein said plurality of genes comprises all known unique genes of said organism other than said target gene.

64. A library of siRNAs, said library comprising a plurality of siRNAs for each of a plurality of different genes of an organism, wherein each siRNA achieves at least 75%, at least 80%, or at least 90% silencing of its target gene.

65. The library of claim 64, wherein said plurality of siNRAs consists of at least 3, at least 5, or at least 10 siRNAs.

66. The library of claim 65, wherein said plurality of different genes consists of at least 10, at least 100, at least 500, at least 1,000, at least 10,000, or at least 30,000 different genes.

67. A method for determining a base composition position-specific score matrix (bsPSSM) $\{\log(e_{ij}/p_{ij})\}$ for representing base composition patterns of siRNA functional sequence motifs of L nucleotides in transcripts, wherein $i = G, C, A, U(T)$ and $j = 1, 2, ..., L$, and wherein each said siRNA functional sequence motif comprises at least a portion of the target sequence of the corresponding targeting siRNA and/or a sequence in a sequence region flanking said target sequence, said method comprising

(a) identifying a plurality of N different siRNAs consisting of siRNAs having a silencing efficacy above a chosen threshold;

(b) identifying a plurality of N corresponding siRNA functional sequence motifs, one for each said different siRNA;

(c) calculating a frequency matrix $\{f_{ij}\}$, where $i = G, C, A, U(T); j = 1, 2, ..., L$, and where f_{ij} is the frequency of the ith nucleotide at the jth position, based on said plurality of N siRNAs functional sequence motifs according to equation

$$f_{ij} = \sum_{k=1}^{N} \delta_{ik}(j),$$

where $\delta_{ik}(j) = \begin{cases} 1, \text{if } k = i \\ 0, \text{if } k \neq i \end{cases}$, and

(d) determining said psPSSM by calculating e_{ij} according to equation

$$e_{ij} = \frac{f_{ij}}{N}.$$

68. The method of claim 67, wherein each said siRNA functional motif comprises the target sequence of the corresponding targeting siRNA and one or both flanking sequences of said target sequence.

69. The method of claim 68, wherein each said siRNA has M nucleotides in its duplex region, and wherein each said siRNA functional sequence motif consists of an siRNA target sequence of M nucleotides, a 5' flanking sequence of D_1 nucleotides and a 3' flanking sequence of D_2 nucleotides.

70. The method of claim 69, wherein each said siRNA has 19 nucleotides in its duplex region, and wherein each said siRNA functional sequence motif consists of an siRNA target sequence of 19 nucleotides, a 5' flanking sequence of 10 nucleotides and a 3' flanking sequence of 10 nucleotides.

71. The method of claim 69, wherein each said siRNA has 19 nucleotides in its duplex region, and wherein each said siRNA functional sequence motif consists of an siRNA target sequence of 19 nucleotides, a 5' flanking sequence of 50 nucleotides and a 3' flanking sequence of 50 nucleotides.

72. The method of claim 67, wherein said plurality of N siRNAs each targets a gene whose transcript abundance is within a given range in a cell.

73. The method of claim 72, wherein said range is at least about 5, 10, or 100 transcripts per cell.

74. The method of claim 72, wherein said range is less than about 3-5 transcripts per cell.

75. The method of any one of claims 67-74, wherein said threshold is 50%, 75%, or 90% at an siRNA dose of about 100nM.

76. The method of any one of claims 67-74, wherein said plurality of N siRNAs comprises 10, 50, 100, 200, or 500 different siRNAs.

77. A method for determining a base composition position-specific score matrix

(bsPSSM) $\{w_{ij}\}$ for representing a base composition pattern representing a plurality of

different siRNA functional sequence motifs of L nucleotides, wherein i = G, C, A, U(T) and j

= 1, 2, ..., L, and wherein each said siRNA functional sequence motif comprises at least a

5   portion of the target sequence of the corresponding targeting siRNA and/or a sequence in a

sequence region flanking said siRNA target sequence, said method comprising

(a) initializing said bsPSSM with random weights;

(b) selecting randomly a weight w_{ij} obtained in (a);

(c) changing the value of said selected weight to generate a test psPSSM comprising

10   said selected weight having said changed value;

(d) calculating a score for each of said plurality of siRNAs functional sequence motifs

using said test psPSSM according to equation

$$Score = \sum_{k=1}^{L} \ln(w_k / p_k)$$

wherein said w_k and p_k are respectively weights of a nucleotide at position k in said functional

15   sequence motif and in a random sequence;

(e) calculating correlation of said score and a metric characterizing an siRNA among

said plurality of siRNAs functional sequence motifs;

(f) repeating steps (c)-(e) for a plurality of different values of said selected weight in a

given range and retain the value that corresponds to the best correlation for said selected

20   weight; and

(g) repeating steps (b)-(f) for a chosen number of times; thereby determining said

psPSSM.

78. A method for determining a base composition position-specific score matrix

(bsPSSM) $\{w_{ij}\}$ for representing a base composition pattern representing a plurality of

25   different siRNA functional sequence motifs of L nucleotides, wherein i = G/C, A, U(T) and j

= 1, 2, ..., L, and wherein each said siRNA functional sequence motif comprises a least a

116

I00005189

portion of the target sequence of the corresponding siRNA and/or a sequence in a sequence region flanking said siRNA target sequence, said method comprising

(a) initializing said bsPSSM with random weights;

(b) randomly selecting a weight w_{ij} obtained in (a);

5   (c) changing the value of said selected weight to generate a test psPSSM comprising said selected weight having said changed value;

(d) calculating a score for each of said plurality of siRNA functional sequence motifs using said test psPSSM according to equation

$$Score = \sum_{j=1}^{L} \ln(w_k / p_k)$$

10   wherein said w_k and p_k are respectively weights of a nucleotide at position k in said functional sequence motif and in a random sequence;

(e) calculating a correlation of said score and a metric of a characteristic of an siRNA among said plurality of siRNAs functional sequence motifs;

(f) repeating steps (c)-(e) for a plurality of different values of said selected weight in a

15   given range and retain the value that corresponds to the best correlation for said selected weight; and

(g) repeating steps (b)-(f) for a chosen number of times; thereby determining said psPSSM.

79. The method of claim 77 or 78, wherein each said siRNA functional motif

20   comprises the target sequence of the corresponding targeting siRNA and one or both flanking sequences of said target sequence.

80. The method of claim 79, further comprising selecting said plurality of siRNA functional sequence motifs by a method comprising

(i) identifying a plurality of siRNAs consisting of siRNAs having different values in

25   said metric;

117

(ii) identifying a plurality of siRNA functional sequence motifs each corresponding to an siRNA in said plurality of siRNAs.

81. The method of claim 79, wherein each said siRNA has M nucleotides in its duplex region, and wherein each said siRNA functional sequence motif consists of an siRNA target sequence of M nucleotides and a D_1 nucleotide flanking sequence upstream and a D_1 nucleotide flanking sequence downstream.

82. The method of claim 81, wherein each said siRNA has 19 nucleotides in its duplex region, and wherein said siRNA functional sequence motif consists of an siRNA target sequence of 19 nucleotides and a 10 nucleotide flanking sequence upstream and a 10 nucleotide flanking sequence downstream.

83. The method of claim 82, wherein each said siRNA has 19 nucleotides in its duplex region, and wherein said siRNA functional sequence motif consists of an siRNA target sequence of 19 nucleotides and a 50 nucleotide flanking sequence upstream and a 50 nucleotide flanking sequence downstream.

84. The method of claim 82, wherein said metric is silencing efficacy.

85. The method of claim 84, wherein said plurality of siRNAs consisting of siRNAs targeting genes whose transcript abundance is in a given range in a cell.

86. The method of claim 85, wherein said range is at least about 5, 10, or 100 transcripts per cell.

87. The method of claim 85, wherein said range is less than about 3-5 transcripts per cell.

88. The method of any one of claims 77-78, wherein said threshold is 50%, 75%, or 90% at an siRNA dose of about 100nM.

89. The method of claim 84, further comprising evaluating said psPSSM using an ROC (receiver operating characteristic) curve of the sensitivity of said psPSSM vs. the non-specificity of said psPSSM curve, said sensitivity of said PSSM being the proportion of true positives detected using said psPSSM as a fraction of total true positives, and said non-

specificity of said PSSM being the proportion of false positives detected using said psPSSM as a fraction of total false positives.

90. The method of claim 84, wherein said plurality of siRNA functional sequence motifs consists of at least 50, at least 100, or at least 200 different siRNAs functional sequence motifs.

91. The method of claim 84, further comprising testing said psPSSM using another plurality of siRNA functional sequence motifs.

92. A method for determining a position match position-specific score matrix (pmPSSM) $\{E_i\}$ for representing position match pattern of an siRNA of L nucleotides with its target sequence in a transcript, wherein E_i is a score of a match at position i, $i = 1, 2, ..., L$, said method comprising

(a) identifying a plurality of N siRNA off-target sequences, wherein each said off-target sequence is a sequence on which said siRNA exhibits silencing activity;

(b) calculating a position match weight matrix $\{P_i\}$, where $i = 1, 2, ..., L$, based on said plurality of N siRNAs off-target sequences according to equation

$$P_i = \frac{1}{N}\sum_{k=1}^{N}\delta_k(j),$$

where $\delta_k(j)$ is 1 if k is a match, and is 0 if k is a mismatch; and

(c) determining said psPSSM by calculating E_i such that $E_i = P_i$ if position i is a match and $E_i = (1-P_i)/3$ if position i is a mismatch.

20   93. The method of claim 92, wherein $L = 19$.

94. The method of claim 93, wherein said position match weight matrix is given by Table I.

95. A method for evaluating the relative activity of the two strands of an siRNA in off-target gene silencing, comprising comparing position specific base composition of the sense strand of said siRNA and position specific base composition of the antisense strand of

119

said siRNA or reverse complement strand of said sense strand of said siRNA, wherein said antisense strand is the guiding strand for targeting the intended target sequence.

96. The method of claim 95, wherein said comparing is carried out by a method comprising

5   (a) determining a score for said sense strand of said siRNA, wherein said score is calculated using a position-specific score matrix;

        (b) determining a score for said antisense strand of said siRNA or said reverse complement strand of said sense strand of said siRNA using said position-specific score matrix; and

10   (c) comparing said score for said sense strand and said score for said antisense strand or said reverse complement strand of said sense strand, thereby evaluating strand preference of said siRNA.

97. The method of claim 96, wherein said siRNA has a nucleotide sequence of L nucleotides in its duplex region, L being an integer, wherein said position-specific score

15   matrix is $\{w_{ij}\}$, where w_{ij} is the weight of nucleotide i at position j, $i = G, C, A, U(T)$, $j = 1$, ..., L.

98. The method of claim 96, wherein said siRNA has a nucleotide sequence of L nucleotides in its duplex region, L being an integer, and wherein said position-specific score matrix is $\{w_{ij}\}$, where w_{ij} is the weight of nucleotide i at position j, $i = G$ or $C, A, U(T)$, $j = 1$,

20   ..., L.

99. The method of claim 97 or 98, wherein said position-specific score matrix is obtained by a method comprising

        (a) initializing said position-specific score matrix with random weights;

        (b) selecting randomly a weight w_{ij} obtained in (a);

25   (c) changing the value of said selected weight to generate a test position-specific score matrix comprising said selected weight having said changed value;

120

(d) calculating a score for each of a plurality of siRNAs using said test position-specific score matrix according to equation

$$Score = \sum_{j=1}^{L} \ln(w_j / p_j)$$

wherein said w_j and p_j are respectively weights of a nucleotide at position j in said siRNA and

5   in a random sequence;

(e) calculating correlation of said score with a metric of a characteristic of an siRNA among said plurality of siRNAs;

(f) repeating steps (c)-(e) for a plurality of different values of said selected weight in a given range and retain the value that corresponds to the best correlation for said selected

10   weight; and

(g) repeating steps (b)-(f) for a chosen number of times; thereby determining said position-specific score matrix.

100. The method of claim 99, wherein said metric is siRNA silencing efficiency.

101. The method of claim 100, wherein said siRNA has 19 nucleotides in its duplex

15   region.

102. The method of claim 96, wherein said siRNA has a nucleotide sequence of L nucleotides in its duplex region, L being an integer, wherein said position-specific score matrix comprises w_k, $k = 1, ..., L$, w_k being a difference in probability of finding nucleotide G or C at sequence position k between a first type of siRNA and a second type of siRNA, and

20   wherein said score for each said strand is calculated according to equation

$$Score = \sum_{k=1}^{L} w_k .$$

103. The method of claim 102, wherein said first type of siRNA consists of one or more siRNAs having silencing efficacy no less than a first threshold and said second type of siRNA consists of one or more siRNAs having silencing efficacy less than a second

25   threshold, and wherein said siRNA is determined as having antisense preference if said score determined in step (a) is greater than said score determined in step (b), or as having sense

121

I00005189

preference if said score determined in step (b) is greater than said score determined in step (a).

104. The method of claim 103, wherein said difference in probability is described by a sum of Gaussian curves, each of said Gaussian curves representing the difference in probability of finding a G or C at a different sequence position .

105. The method of claim 104, wherein said first and second threshold are both 75% at an siRNA dose of about 100nM.

106. A computer system comprising
   a processor, and
   a memory coupled to said processor and encoding one or more programs, wherein said one or more programs cause the processor to carry out the method of any one of claims 1-20, 28, 48-50, 52-56, 67-71, 77-98, and 104.

107. A computer system comprising
   a processor, and
   a memory coupled to said processor and encoding one or more programs, wherein said one or more programs cause the processor to carry out the method of claim 26.

108. A computer system comprising
   a processor, and
   a memory coupled to said processor and encoding one or more programs, wherein said one or more programs cause the processor to carry out the method of claim 27.

109. A computer system comprising
   a processor, and
   a memory coupled to said processor and encoding one or more programs, wherein said one or more programs cause the processor to carry out the method of claim 37.

110. A computer program product for use in conjunction with a computer having a processor and a memory connected to the processor, said computer program product comprising a computer readable storage medium having a computer program mechanism encoded thereon, wherein said computer program mechanism may be loaded into the memory

of said computer and cause said computer to carry out the method of any one of claims 1-20, 28, 48-50, 52-56, 67-71, 77-98, and 104.

111. A computer program product for use in conjunction with a computer having a processor and a memory connected to the processor, said computer program product

5   comprising a computer readable storage medium having a computer program mechanism encoded thereon, wherein said computer program mechanism may be loaded into the memory of said computer and cause said computer to carry out the method of claim 26.

112. A computer program product for use in conjunction with a computer having a processor and a memory connected to the processor, said computer program product

10 comprising a computer readable storage medium having a computer program mechanism encoded thereon, wherein said computer program mechanism may be loaded into the memory of said computer and cause said computer to carry out the method of any one of claim 27.

113. A computer program product for use in conjunction with a computer having a processor and a memory connected to the processor, said computer program product

15 comprising a computer readable storage medium having a computer program mechanism encoded thereon, wherein said computer program mechanism may be loaded into the memory of said computer and cause said computer to carry out the method of claim 37.